DTU Library

Timbral aspects of reproduced sound in small rooms. II

Bech, Søren

Published in:
Acoustical Society of America. Journal

Link to article, DOI:
10.1121/1.414952

Publication date:
1996

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Bech, S. (1996). Timbral aspects of reproduced sound in small rooms. II. *Acoustical Society of America. Journal*, *99*(6), 3539-3550. https://doi.org/10.1121/1.414952

Timbral aspects of reproduced sound in small rooms. II

Søren Bech[a)]
The Acoustics Laboratory, Technical University of Denmark, DK-2800 Lyngby, Denmark

(Received 25 May 1995; accepted for publication 30 January 1996)

This paper is a report on some of the influences of individual reflections on the timbre of reproduced sound. Bech [J. Acoust. Soc. Am. **97**, 1717–1726 (1995)] gave the first report. A single loudspeaker with frequency-dependent directivity characteristics, positioned in a room of normal size with frequency-dependent absorption coefficients of the room surfaces, has been simulated using an electroacoustic setup. The model included the direct sound, 17 individual reflections and the reverberant field. The threshold of detection, and just-noticeable differences for an increase in level were measured for individual reflections, using four subjects for noise and three for speech. The results have confirmed the findings of the first report that the first-order floor reflection is likely to individually contribute to the timbre of reproduced noise. However, for a speech signal none of the investigated reflections will contribute individually to the timbre. It is suggested that the threshold of detection is determined by the spectral changes in the dominant frequency range of 500 Hz–2 kHz. For increases in the level of individual reflections, the most likely to be audible is the first-order floor reflection, for speech and noise. For a noise signal, additional reflections from the wall to the left and behind the listener also belong to this group.

PACS numbers: 43.55.Hy, 43.55.Jz, 43.66.Jh

INTRODUCTION

This paper is the second report on the psychoacoustic results of the Archimedes project. The first set of results were discussed in Bech.[1] The purpose of the experiments in the first and the present report is to examine the influence of individual reflections on the timbre of sound reproduced by a single loudspeaker in a domestic listening room. To facilitate the investigations, the sound field from the right-hand loudspeaker of a stereophonic setup in a listening room of normal size has been simulated using an electroacoustic setup. Two basic questions have been investigated in the project:

(1) Which early reflections are sufficiently strong to contribute individually to overall timbre, and which only contribute collectively?

(2) How much must the level of an individual reflection change to produce a change in the overall timbre of the sound field?

The results reported in the first report were based on an early version of the simulation setup, in which the transfer function of individual reflections only included frequency independent attenuation. The results described here are based on an improved version of the setup, in which the transfer function of individual reflections in addition to the attenuation due to distance, also included the frequency response of the off-axis angle of the reflection path from the original loudspeaker, and the frequency-dependent attenuation of the reflection from the simulated room surfaces. Therefore, subjective results presented in the following should be closer to the results that would have been found in a real room.

The results are based on the same basic experimental setup, experimental procedure, group of subjects, and stimuli as used in the first report so in general the reader is referred to this report. However to facilitate the reading, a short introduction to the experimental setup will be given in the following plus to the small number of changes that have been made. The paper is organized as follows: Secs. I–IV contain a short description of the setup, subjects, and the general procedure. Section V contains the results and the discussion, Sec. VI a general discussion, and Sec. VII a summary of the findings.

The Archimedes project was a joint effort between Bang and Olufsen (DK), KEF Electronics Ltd. (GB) and The Acoustics Laboratory of The Technical University of Denmark. The project has been partially financed under the European research program, EUREKA.

I. EXPERIMENTAL SETUP

The setup models the direct sound, 17 individual reflections arriving less than 22 ms after the arrival of the direct sound, and the reverberant part of the sound field or reflections arriving more than 22 ms after the arrival of the direct sound. The setup was positioned in the large (1000 m^3) anechoic chamber of The Acoustics Laboratory, and all loudspeakers were located, with correct azimuth and elevation, on the surface of an imaginary sphere of 3-m radius centered on the listening position. The positions of all the loudspeakers, and the delay and attenuation of all signals representing individual images and reverberation channels, are given in Table I. In the following, individual reflections will be identified either by the delay *re*: the direct sound, or by the number given in Table I.

[a)]Present address: Bang and Olufsen A/S, Peter Bangs Vej 15, DK-7600 Struer, Denmark.

TABLE I. Position of loudspeakers and delay and attenuation of the signals to the loudspeaker for primary loudspeaker and images and reverberation channels included in the setup. The attenuation values are only based on attenuation due to distance. The last wall of the reflection path is also given. All angles and wall references are relative to the listening position. The left-hand side of the subject defines positive angles.

Delay [ms]	Attenuation [dB]	Azimuth [degrees]	Elevation [degrees]	Reflection number	Last surface of reflection
0	0	−22	0	⋯	primary lsp
1.64	1.36	−25	−28	1	floor
4.16	3.1	−50	−2	2	right wall
4.48	3.28	−25	48.2	3	ceiling
5.36	3.81	−53	−28	4	floor
7.6	5.01	−50	48	5	ceiling
9.2	5.78	−25	48.2	6	ceiling
9.2	5.78	−25	−56	7	floor
9.94	6.11	65	0	8	left wall
10.8	6.48	65	−14	9	left wall
11.64	6.83	−53	−56	10	floor
11.64	6.83	−50	48	11	ceiling
12.5	7.17	65	30	12	left wall
12.7	7.25	−170	0	13	back wall
13.46	7.54	−170	−15	14	back wall
14.42	7.9	−25	−56	15	floor
14.8	8.03	−154	0	16	back wall
14.98	8.09	−170	33	17	back wall
22	6	71	0	⋯	rev. syst.
22	6	−71	0	⋯	rev. syst.
22	6	127	0	⋯	rev. syst.
22	6	−127	0	⋯	rev. syst.
22	6	180	0	⋯	rev. syst.
22	6	0	0	⋯	rev. syst.

A. Implementation of the direct sound and individual reflections

The modeled loudspeaker was a two-way system (KEF 103.2) with an 8-in. woofer and a 1-in. tweeter, and a crossover frequency of 2.5 kHz. They were mounted in a closed box of dimensions (w×h×d) 264×501×240 mm. The free-field frequency response of the loudspeaker was measured in directions corresponding to the position of the images given in Table I at a distance of 3 m, with the front cover removed. The geometrical center of the baffle front was defined as the center of the loudspeaker.

The frequency-dependent absorption of the room surfaces was modeled according to measurements of the diffuse field absorption coefficient and the cosine law[2] in the following way.

The absorption material used on the walls in the modeled listening room was distributed in such a way that the same mean absorption coefficient could be used for all four walls, and was estimated based on diffuse field measurements of the individual components. The absorption coefficients for the floor and the ceiling were also based on diffuse field measurements.

The absorption coefficients for the walls, the floor, and the ceiling are given in Table II. The absorption coefficient as a function of angle was found by setting the diffuse field coefficient equal to the absorption at an angle-of-incidence of 45 deg, and then applying the cosine law for other angles. Rindel[3] discusses the derivation of an angle-dependent absorption coefficient based on diffuse field measurements.[4]

The frequency responses of the signal paths to the individual loudspeakers in the simulation setup were calculated taking into account the directivity characteristics and absorption coefficients as discussed above, and were implemented as digital filters.[5–7]

The implemented transfer function for selected reflections are shown in Fig. 1. Note that the transfer functions have been adjusted to include the attenuation due to distance, as given in Table I.

B. Implementation of the reverberant field

The experience obtained during the experiments described in the first report suggested that the subjective diffuseness of the reverberant field could be improved by changing the method of its simulation. This was done by reducing the correlation between the six individual channels that created the reverberant field. The original setup included six loudspeakers, positioned in the equatorial plane of the imaginary sphere described above. Signals for the six loudspeakers were based on the two uncorrelated outputs from a commercially available reverberation unit (Lexicon PCM70). To reduce the correlation between the six loudspeaker signals in the new setup, two reverberation units were added and the settings of the three PCM70 units were set slightly differently. This provided six uncorrelated signals which were fed directly to the six loudspeakers. A block diagram of the complete system is shown in Fig. 2.

The level of the reverberant part of the sound field, relative to the direct sound and early reflections, was adjusted so that it corresponds to the ratio measured at 1 kHz in the

TABLE II. Diffuse field absorption coefficients for the various room surfaces as a function of one-third octave frequencies. Note that the coefficients for the walls, floor, and ceiling are assumed to be constant and at the level of 50 Hz and 8 kHz, respectively, for frequencies outside the 50–8000 Hz range.

One-third oct. frequency [Hz]	Absorption coefficient for walls	Absorption coefficient for floor	Absorption coefficient for ceiling
50	0.05	0.05	0.15
63	0.17	0.06	0.13
80	0.28	0.07	0.11
100	0.45	0.08	0.1
125	0.46	0.09	0.09
160	0.35	0.1	0.08
200	0.34	0.12	0.08
250	0.41	0.14	0.07
315	0.37	0.16	0.07
400	0.4	0.19	0.07
500	0.41	0.24	0.06
630	0.33	0.28	0.06
800	0.25	0.33	0.06
1000	0.24	0.35	0.05
1250	0.31	0.33	0.05
1600	0.15	0.31	0.05
2000	0.16	0.28	0.04
2500	0.18	0.25	0.04
3150	0.16	0.22	0.04
4000	0.14	0.2	0.03
5000	0.18	0.18	0.03
6300	0.18	0.16	0.03
8000	0.19	0.14	0.02

listening room being simulated. The ratio of the level of the direct sound and early reflections to the level of the reverberant field is given as a function of frequency for the real room and the simulation setup in Table III. Reverberation time and the timbral character of the reverberant field was not changed by introducing the two additional reverberation units.

C. Subject positioning and calibration procedures

The listener's ears were moved to the specified listening position using a motorized adjustment mechanism built into the chair supporting the subject, and a fixed video camera. A curtain prevented the listener from seeing the simulation setup, while a single LED was used to define the front angular reference. Listeners were free to move their heads, but were instructed to focus attention on the LED. The performance and calibration of the entire setup was checked on a daily basis using a PC-controlled measuring system.

The reproduction level, measured at the listening position with a single microphone, was 66 dB SPL for the noise stimulus, and approximately 50 dB SPL (time weighting fast) for the speech stimulus. The background noise level, with the setup operating, was 27 dB SPL (time weighting fast) with the one-third octave levels constant at ± 2 dB, for the frequency range 20 Hz–20 kHz.

II. STIMULI

Broadband pink noise and speech were used a representatives of continuous and discontinuous sounds, respectively. The signals were identical to those used in the first report. The noise signal was a 1-s sample of broadband (20 Hz–20 kHz) pink noise, and the speech signal was a 3.8-s sample of male speech. The speech was an anechoic recording of an excerpt of the text used for the standardized Danish speech material for audiometric purposes, and its time structure and spectrum resembled average Danish speech. For the investigations of a dominant frequency range, three high-pass filtered (24 dB/oct at 500 Hz, 1 kHz, or 2 kHz) and three low-pass filtered (24 dB/oct at 500 Hz, 1 kHz, or 2 kHz) versions of the noise signal were used.

All signals were digitally stored and played back via 16-bit D/A at a sampling rate of 50 kHz, with low-pass filtering at 20 kHz. The rise and fall time of the noise signals were 5 ms following a linear function.

III. SUBJECTS

The subjects were paid an hourly rate for participating in the experiments. Each subject had received a total of approximately 40 000 trials before participating in the experiments reported here. The subjects were divided into two groups; a group of four subjects who worked only with noise signals, and a group of three who worked only with speech signals. Before participating in the main experiments, the subjects in each group participated in two training experiments, which included a total of 800 trials for noise signals and 600 trials for speech signals. The same training experiment has been used at regular intervals throughout the whole project as a check on the performance of each subject. See Bech[8] for a discussion of the training experiments.

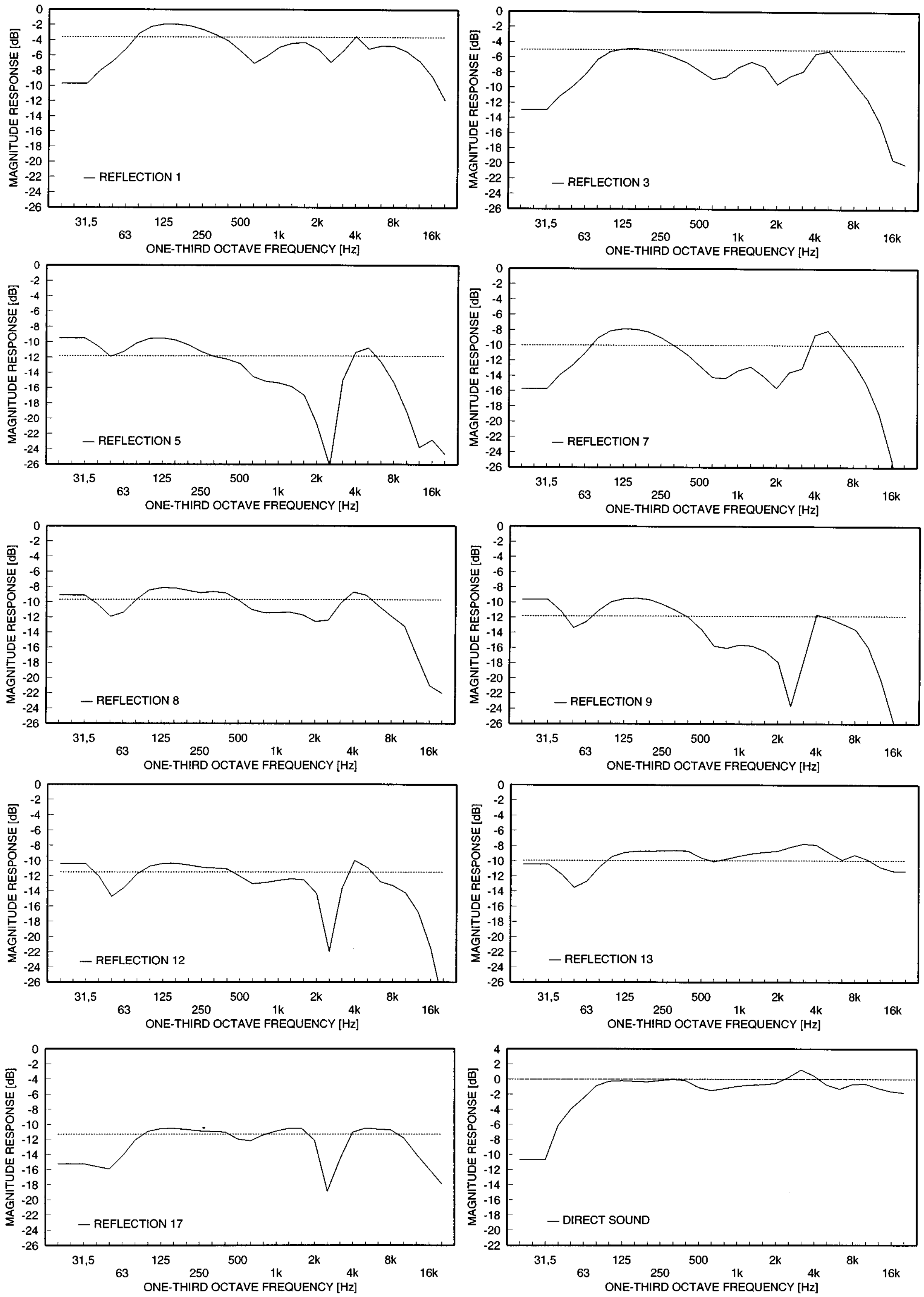


FIG. 1. Magnitude response of the filter function implemented for selected individual reflections (solid line). The dashed line represents the frequency independent attenuation used in Bech.[1] The reflection numbers refer to Table I.

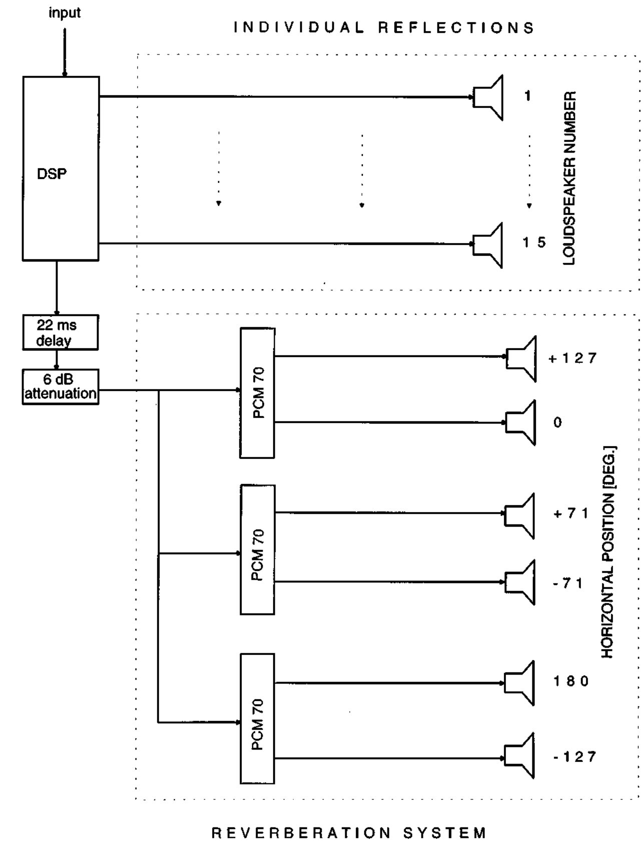


FIG. 2. Block diagram of the complete experimental setup. The DSP unit implemented delay and attenuation due to distance and the transfer functions shown in Fig. 1. Note that the reflection loudspeakers can represent more than one image, cf. Table I and that 0 deg corresponds to the front angular reference for the subject and positive angles are to the left-hand side.

IV. GENERAL PROCEDURE

The task of the subjects in all the experiments was to detect a change in timbre of a pink noise signal or a speech stimulus. The interpretation of timbre[9] was discussed with the subjects during the entire experimental period to ensure that their understanding remained constant. For each of the reflections Nos. 1, 5, 7, 9, 13, and 17,[10] two psychoacoustic quantities were determined (a) threshold of detection (TD) and (b) just noticeable difference (jnd) corresponding to questions 1 and 2, respectively, as asked for in the introduction. An adaptive (staircase) two-alternative forced-choice procedure was used. See the first report for a detailed description of this procedure. The standard and comparison stimuli for the two situations are defined as follows:

(1) Threshold of detection (TD)

The **standard** is the complete sound field simulating a loudspeaker in the listening room, except that the reflection under investigation is absent (i.e., attenuated 100 dB *re*: direct sound). The **comparison** stimulus was formed by adding a variable level of the reflection under investigation to the standard.

(2) Just-noticeable difference (jnd)

The **standard** was the complete sound field simulating a loudspeaker in the listening room. The **comparison** stimulus is derived from the standard by a variable increase in the level of the reflection under investigation.

In the TD experiments, the initial level of the reflection under investigation was equal to the level of the direct sound. For the jnd experiments, the initial level was a 10-dB increase in the level of the reflection under investigation. The level of the reflection under investigation was varied adaptively (two down/one up) to estimate the level that would produce 70.7% correct responses (Levitt[11]). The step size was initially 4 dB, reduced to 2 dB (absolute threshold experiments), or 1 dB (jnd experiments) after three reversals. Typically 10–15 reversals would occur during each 50-trial block. For each block the threshold was estimated as the average of the midpoints of runs 4, 6, 8, etc. The reported TD or jnd was averaged across subjects for eight (noise) or six (speech) block estimates per subject. The comparison stimulus occurred with equal probability in the first or second period. The other period contained the standard. The two observation periods were separated by a 0.5-s silent interval.

Other details of the experimental procedure can be found in the first report, and an overview of the present series of experiments can be found in Table IV.

TABLE III. The level of the direct sound and the early reflections relative to the level of the diffuse part of the sound field as measured in the real room and in the simulation setup.

One octave frequency [Hz]	Ratio for real room [dB]	Ratio for simulation [dB]
125	−0.8	2.2
250	0.4	5
500	4.6	5.2
1000	5.3	5.3
2000	6.4	5.4
4000	4.4	9.1
8000	6.5	11.7

V. RESULTS AND DISCUSSION

A. Threshold of detection for individual reflections

The purpose of these experiments (Nos. I–III in Table IV) was to measure the threshold of detection (TD) for individual reflections under different conditions, and to compare the TD values with the natural level of the reflection in a standard listening room. The experiments are related to question 1 in the introduction.

1. Comparison of natural levels and measured thresholds of individual reflections

The threshold of detection for noise and speech signals are shown in Fig. 3, together with the natural levels. The natural levels have been defined to be the same as those used in Bech[1] and given as the dashed lines in Fig. 1. This is a compromise to describe the transfer functions given in Fig. 1 by a single number. It is seen to be a reasonable approximation for reflections Nos. 1, 13, and 17. For reflections Nos. 5, 7, and 9 the approximation is about 4 dB too high, as the discussion to follow in Sec. V A 2 indicates that the frequency region 500–2 kHz is the most important for the threshold of detection.

The general tendency is for the natural level to be lower than the TD's, with the only exception being reflection 1 for

TABLE IV. Overview of the experiments that are discussed in this paper. The reflection numbers refer to Table I.

Exp. No.	Description of experiment	Stimulus	Filtering	No. of subjects	Results in Fig.
	Measurement of TD for reflection number:				
I	* 1,5,7,9,13,17	noise	yes	4	3
	* 1,5,7,9,13,17	speech	yes	3	3
	Measurement of TD for reflection number:				
II	* 1,5,7,9,13,17	noise	no	4	4
	* 1,7,13,17	speech	no	3	5
	Measurement of TD for reflection number:				
III	* 1	lp&hp noise	no	4	6
	* 9	lp&hp noise	no	4	6
	Measurement of jnd for reflection number:				
IV	* 1,5,7,9,13,17	noise	yes	4	7
	* 1,5,7,9,13,17	speech	yes	3	7

the noise signal. This is in agreement with previous results[1] which showed that only reflections 1, 3, 8, and 12 were either significantly lower, or similar to the natural levels for the noise signal. While the results shown in Fig. 3 have confirmed this for reflection 1, the TD's for reflections Nos. 3, 8, and 12 are not available. However, the discussion in Sec. V A 2 will show that the TD's for those reflections are likely to be lower, or at the natural levels, in accordance with the findings of the first report. The speech signal is seen to result in a significant increase in the TD values. This has been observed previously,[1] but the increase was generally smaller. The results shown in Fig. 3, therefore, confirm the findings of the first report, that the floor reflection (No. 1) will contribute to the timbre of the sound field on an individual basis, for a noise signal.

2. The effects of filtering the transfer function of the individual reflections

To examine the influence of the transfer functions that were introduced for the individual reflections, the TD's for reflections Nos. 1, 5, 7, 9, 13, and 17 were measured without filtering, for noise and speech. For this experiment, the level of the reverberant part of the sound field relative to the direct sound and early individual reflections, was adjusted for the **new** system to be identical to that of the system used in Bech.[1]

The results in Fig. 4 for the noise signal, and in Fig. 5 for the speech signal, show that the filtering introduced only influences the TD values for the noise signals as the TD's for reflections Nos. 5, 7, and 9 increase significantly. A visual inspection of the transfer functions given in Fig. 1 shows that these reflections have differences of 4 dB or more between

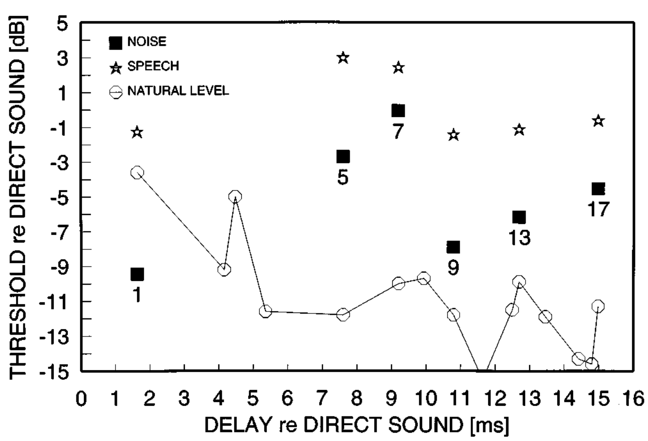


FIG. 3. Threshold of detection for the noise and speech signals for reflections Nos. 1, 5, 7, 9, 13, and 17 (experiment I). The natural levels (see text) of the individual reflections are also shown. The reflection numbers refer to Table I. Confidence intervals (95%) are ±0.94 dB for both speech and noise. The confidence intervals are based on the variance between blocks and mean values are based on four subjects and 400 trials per subject for noise and three subjects and 300 trials for the speech signal.

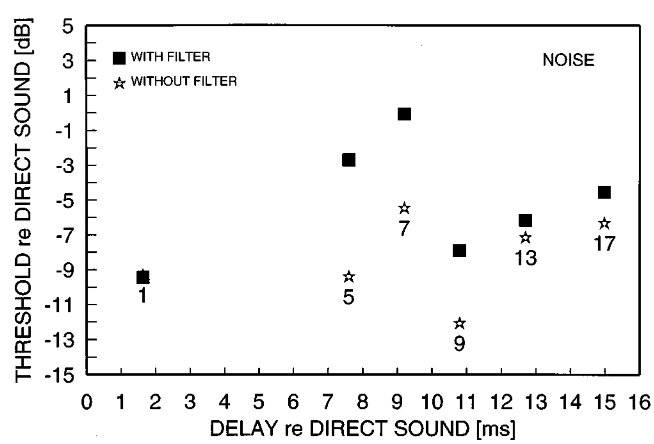


FIG. 4. Threshold of detection for the noise signal for reflections Nos. 1, 5, 7, 9, 13, and 17 with filtering (experiment I) and without (experiment II). Confidence intervals (95%) are ±0.94 dB with filtering and ±0.96 dB without filtering. The confidence intervals are based on the variance between blocks and the mean values are based on the same four subjects for both experiments and 400 trials per subject.

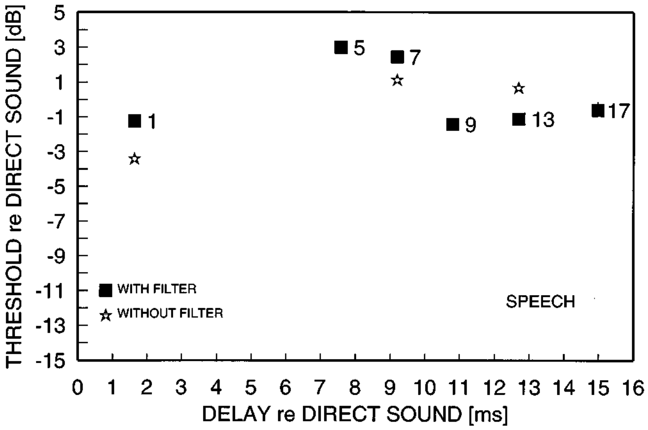
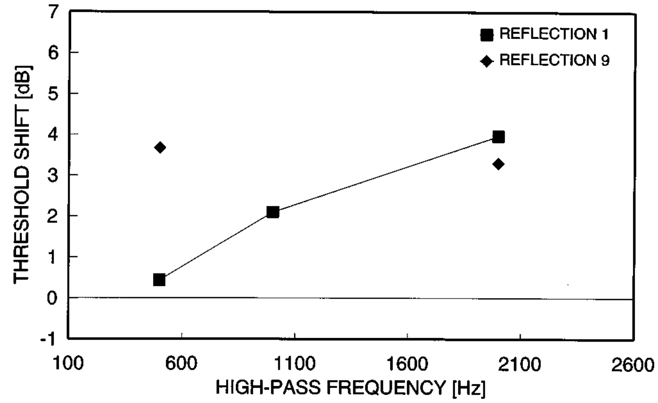


FIG. 5. Threshold of detection for the speech signal for reflections Nos. 1, 5, 7, 9, 13, and 17 with filtering (experiment I) and reflections 1, 7, 13, and 17 without (experiment II). Confidence intervals (95%) are ± 0.94 dB with filtering and ± 1.07 dB without filtering. The confidence intervals are based on the variance between blocks and the mean values are based on the same three subjects for both experiments and 300 trials per subject.

the filtered and unfiltered transfer functions in the frequency regions 500 Hz–2 kHz and above 5–6 kHz. However, such differences are not seen for reflections 1, 13, and 17 which suggests that the significant increase in the TD values for reflections 5, 7, and 9 could be caused by removal of energy in the mid- and high-frequency regions. This would be in agreement with results presented by Olive and Toole.[12] If this assumption holds (for a further discussion see the next section), it follows from the transfer functions of reflections 3, 8, and 12 seen in Fig. 1, that the TD's for these reflections would not be significantly influenced by the introduction of filtering. Previous findings,[1] which showed that the TD's for reflections 3, 8, and 12 were not significantly different from the natural levels, are thus also likely to apply to the situation with filtering.

The results in Fig. 5 show that the filtering introduced had no effect on the TD's for the speech signal. This could be explained by the fact that the speech signal has its main energy in the frequency range 70–500 Hz as the spectral level is 12–15 dB lower in the range 700 Hz–2 kHz compared to 70–500 Hz. The transfer functions in Fig. 1 show that the filtering only introduces relatively small changes in this frequency range, and consequently only small changes in the TD's would be expected. Another explanation is that the TD's for speech could be determined by loudness differences instead of timbre differences. This would also explain why the TD's are independent of the delay time of the reflections. Such independence was not observed in the first report. See Sec. VI A for a further discussion of possible detection cues at TD.

The results presented therefore suggest that the situation investigated in Bech[1] is the most sensitive for the majority of reflections, and that the introduction of acoustically more realistic conditions will reduce the influence of these individual reflections. The results also indicate that the increase observed in threshold values can be explained by the removal of energy in the mid- and high-frequency ranges.



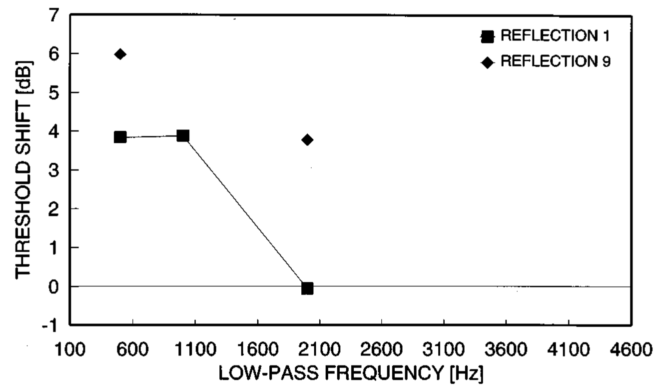


FIG. 6. Shift in threshold of detection as a result of using high-pass filtered (top) or low-pass filtered noise (bottom) compared to broadband noise for reflections Nos. 1 and 9 (experiment III). Positive shifts in TD means that the reflection is less detectable. Confidence intervals (95%) are between ± 0.77 to ± 1.45 dB and they are based on the variance between blocks and the mean values are based on four subjects and 400 trials per subject.

3. Threshold of detection for individual reflections for high- and low-pass filtered noise signals

The purpose of this experiment was to measure the threshold of detection for two selected reflections, using high-pass and low-pass filtered noise, as a function of the cutoff frequency. The experiment was suggested by the results discussed in Sec. V A 2, which indicated that filtering the transfer functions of individual reflections in the form of mid- and high-frequency attenuation, caused the threshold values to increase. The unfiltered version of the transfer functions, shown as dashed lines in Fig. 1, were used for the experiments.

The shift in TD's, as a result of using either low-pass or high-pass filtered noise, is shown in Fig. 6 for reflections Nos. 1 and 9. The tendency is that the threshold increases for increasing cutoff frequency for high-pass filtered signals, and decreases for low-pass filtered signals. The results for reflection 1 suggest that the spectral changes in the frequency range 500 Hz–2 kHz determine the TD for that reflection. The results for reflection 9 are less conclusive, although they do indicate that the frequency range should be extended to approximately 4 kHz. Others[12–17] have also presented evidence of spectrally dominant frequency ranges in the formation of timbral differences. Olive and Toole[12] found that the threshold of detection increased significantly when they low-pass filtered a broadband noise signal. Bilsen and Ritsma[13] showed that the dominant frequency range has a bandwidth of between one-third and a whole octave, and that the center frequency is given by $(3.9\pm0.2)/\tau$ [Hz], where τ is the delay in s of the reflection relative to the direct sound. Bilsen and Ritsma[18] also showed that reflections with a dominant frequency range between 800–1600 Hz will have the lowest TD values. Bismark[15,16] found that sharpness, a main attribute of timbre, is determined by the center of gravity of specific loudness. The center of specific loudness for the present signals is approximately 12 Bark or 1600 Hz. The profile analysis theory[17] has indicated that the sensitivity to spectral changes in a standard profile spectrum has a bowl shaped form, with maximum around 1 kHz.

The results in Fig. 6 support the notion of a dominant frequency range for the spectral changes that determine timbral differences. The frequency range 500 Hz–2 kHz suggested by the results of reflection 1 is in agreement with results in the literature.

4. The influence of the level of the reverberant field

The level of direct sound and early reflections relative to the level of the reverberant field is approx. 0.5 dB for the ''unfiltered'' simulation, and approximately 5 dB in the frequency range 500 Hz–2 kHz for the ''filtered'' simulation. The first report showed that the TD's will decrease by approximately 4.5 dB if the reverberant field is removed from the sound field. This value was found to be independent of delay time for delays less than 13 ms. This suggests that ''filtered'' TD's would be lower than the ''unfiltered'' due to the higher ratio of direct and early reflections to reverberant energy. The results for reflections 1, 13, and 17 shown in Fig. 4 do not support this assumption, and the differences for reflections 5, 7, and 9 have been accounted for in previous sections. This suggests that the level of direct sound and early reflections relative to the level of the reverberant sound has limited influence on the threshold values for ratios within the range 0.5–5 dB.

B. Just noticeable differences in level for individual reflections

The purpose of this experiment (no. IV in Table IV) was to measure the just noticeable difference (jnd) for an increase in level of selected individual reflections. The individual reflections had transfer functions in accordance with the directivity of the real loudspeaker and absorption of the room surfaces.

1. Results

The positive jnd values for noise and speech signals are shown in Fig. 7. The jnd's for the noise signal are seen to divide the reflections into two groups: one group includes reflections 5, 7, and 17 with jnd's that are not significantly different and another, including reflections 1, 9, and 13 that all have jnd's significantly lower than the reflections in the first group. The jnd's for the unfiltered situation given in the first report also resulted in two groups: reflections 1, 2, 3, 8, and 9 in one group and the other reflections in the other. The grouping seen in Fig. 7 for the noise signal is therefore in agreement with the results reported previously, except for reflection 13.

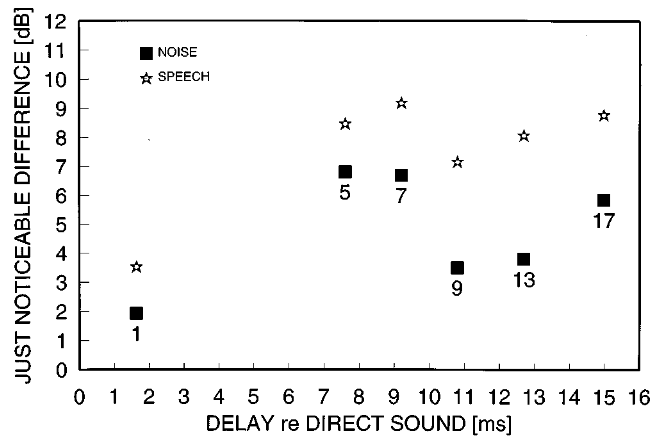


FIG. 7. Just noticeable difference (jnd) for an increase in level for reflections Nos. 1, 5, 7, 9, 13, and 17 for the noise and speech signals (experiment IV). Confidence intervals are ±0.42 dB (noise) and ±0.94 dB (speech). The confidence intervals are based on the variance between blocks and the mean values are based on four subjects and 400 trials per subject (noise) or three subjects and 300 trials per subject (speech).

The jnd's for the speech signal are all significantly higher compared to the noise signal, except for reflection 1. The first report only found a significant increase for reflection 13 and 17 for the speech signal. This confirms the tendency observed for the TD's, that the differences between the noise and speech signals are larger in this experiment than seen before.

Thus the results presented have confirmed earlier findings that the floor reflection (no. 1) will influence the timbre of the sound field to a higher degree than any of the other reflections investigated.

2. Comparison of just noticeable differences for individual reflections with and without filtering

The jnd values for reflections 1, 5, 7, 9, 13, and 17 for the noise and speech signals are shown in Figs. 8 and 9, respectively, together with the results for the unfiltered situation.[19] The results for the noise signal show that the jnd values for reflections 7, 13, and 17 are significantly lower for the filtered reflections than for the unfiltered (note that reflection 7 is only just significantly lower). For the speech signal, only the jnd for reflection 17 is significantly lower for the filtered version. The general tendency for the jnd values for both noise and speech signals is that the introduction of filtering has no influence, except for reflections 7, 13, and 17.

The deviating results for reflections 13 and 17 was also found in a comparison of TD values obtained in the previous sound field and the present field without filtering.[19] They are therefore believed not to be the result of the filtering introduced.

For the speech signal, it is seen that there is no effect of introducing filtering, and this is in agreement with the TD results shown in Fig. 5. As for the TD results, it is probably caused either by the lack of energy in the speech signal in the frequency ranges affected by the filtering, or that detection

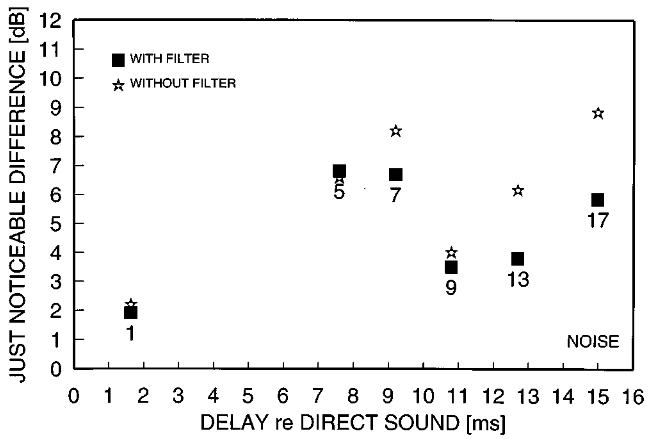


FIG. 8. Just noticeable difference in level for reflections Nos. 1, 5, 7, 9, 13, and 17 with filtering (experiment IV) and without filtering (experiment V in Bech[1]) for noise. Confidence intervals (95%) are ±0.42 dB with filtering and ±0.45 dB without filtering. The confidence intervals are based on the variance between blocks and the mean values are based on four subjects and 400 trials per subject.

was based on a loudness cue (see Sec. VI A). The lack of an effect of filtering for the noise signal, however, is not in agreement with the TD results shown in Fig. 4. Assuming that the TD or jnd is determined by a spectral difference between the standard and the comparison stimuli, this discrepancy could be explained as follows: In the TD experiment the spectral differences between the standard and the comparison stimuli are the comb filter characteristics generated by adding the investigated reflection to the standard. If the filtering removes energy from the reflection in a frequency region that determines the TD, it follows that the level of the reflection must be increased correspondingly to produce the spectral difference needed at TD. For the jnd experiment, the spectral difference will be a change in an already existing comb filter, as the reflection investigated is always present in the standard. Removal of energy in a dominant frequency region of the reflection will thus have a smaller effect, as the spectrum of the standard also changes. Thus it could be expected, and as observed above, that the introduced filtering will have a smaller effect for the jnd values compared to the TD values.

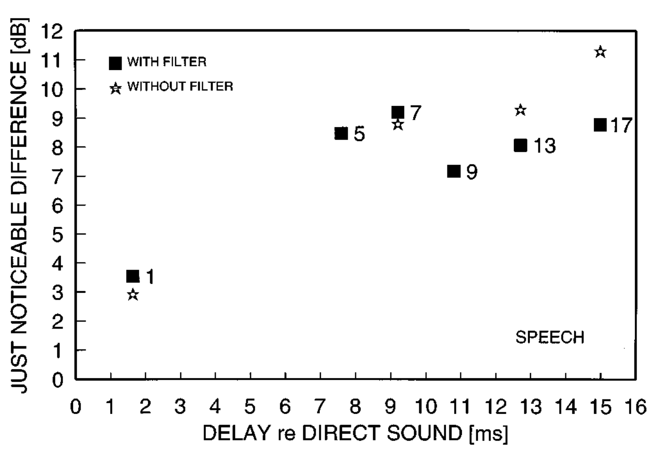


FIG. 9. Just noticeable difference in level for reflections Nos. 1, 5, 7, 9, 13, and 17 with filtering (experiment IV) and reflections Nos. 1, 7, 13, and 17 without filtering (experiment V in Bech[1]) for speech. Confidence intervals (95%) are ±0.94 dB with filtering and ±0.7 dB without filtering. The confidence intervals are based on the variance between blocks and the mean values are based on three subjects and 300 trials per subject.

VI. GENERAL DISCUSSION

A. The detection cues for TD and jnd

The discussion in previous chapters has assumed that the subjects followed the instructions and used timbral differences to establish the TD or jnd. However, this should be verified and to examine the possible use of loudness differences as cues, the SPL differences between the standard and the comparison stimulus, with the reflection at TD or jnd for the noise signal, were measured. The procedure ensured that objective measurement errors would not influence the results. The measured SPL differences were below 0.2 dB for the investigated reflections for the TD results, and below 0.3 dB for the jnd results. A literature study[1] suggests that the jnd for a level difference between two broadband noise signals is in the range 0.5–0.65 dB. Thus the measured values strongly suggest that loudness was not used as a cue for the noise signal in any of the present experiments.

The level difference limen of test material used in speech audiometry is approximately 0.5 dB for an experimental procedure that resembles the present.[20] The level difference limen for speech is thus similar to that for broadband noise. The SPL differences between the standard and the comparison stimulus, with the reflection at TD or jnd for the speech, have not been measured. However, previous measurements for the noise signal, and reflections with TD's at levels similar to those for the speech signal in Fig. 3, suggested that loudness could have been used as a detection cue for those reflections. It can therefore not be excluded that the TD's for the speech signal are based on a loudness cue instead of timbre. This could also explain why the TD's are independent of delay time of the reflection, contrary to the observations of the first report, and also the limited effect of introducing filters to the transfer functions.

B. Generality of the results

Bech[1] concluded when discussing the generality of the results that the realism of the electroacoustic simulation was limited by the fact that the absorption coefficients of the room surfaces were not modeled as a function of frequency. Further, that the directivity of the simulated loudspeaker was modeled as a cordiod, independent of frequency. Thus to improve the realism, the absorption coefficients and the directivity characteristics of a real loudspeaker were implemented as a function of frequency as discussed in Sec. I A. As a further step, to increase the realism, the subjective diffuseness of the reverberant field was improved by increasing the number of uncorrelated channels generating the reverberant part of the sound field as discussed in Sec. I B. The combined effect of these changes was a significant increase in the realism of the simulation. This was verified by researchers with experience in electroacoustic simulation, and there was general agreement about the improvements. The largest improvement was due to the increased diffuseness of the reverberant field.

To further improve the realism of the simulation, it could be considered to include the effect of the power response of the loudspeaker. The directivity characteristics were modeled correctly for the individual reflections, but the frequency response of the reverberant field was only controlled via the limited possibilities for frequency shaping of the reverberation units.

Another point of importance for the generality of the results is the fact that only one combination of loudspeaker-listener positions was modeled. Bech[21] has shown that loudspeaker position has a significant influence on the timbral quality of reproduced sound. However, if it is assumed that changes in the loudspeaker-listener positions always maintain a fixed distance between the loudspeaker and the listener, it follows that the floor and ceiling reflections will have constant delays and levels relative to the direct sound,[22] and changes in the distribution of delays and attenuation's will be only for the other reflections. As both the results of this and the previous report suggest that the floor and ceiling reflections will contribute on an individual basis, it follows that the results obtained will apply for the majority of loudspeaker-listener positions.

A basic limitation of all electroacoustic simulation systems is the missing influence of the standing wave structure of the modeled room. The stationary low-frequency response of the loudspeaker-room-listener system could be modeled by equalizing, but the changes corresponding to the subject moving in the chair are difficult to model. Using a head tracking device connected to a real time equalizer could possibly solve this problem, but such possibilities did not exist at the time of the experiments. The importance of modeling such changes, however, seems to be rather limited in the present situation if the hypothesis of a dominant frequency region between 500 Hz–2 kHz is correct.

To conclude, it is believed that the modified simulation setup used for the present set of experiments is as close as is technically possible, and subjectively needed to model a real situation. The results are therefore believed to be representative of conditions in a real room.

VII. SUMMARY OF FINDINGS

This section contains a summary of the main findings. They are all based on an electroacoustical simulation of the sound field produced by the right-hand loudspeaker of a standard stereophonic setup, positioned in a small room. The validity of the simulation, and correspondingly the results, have been discussed in the previous section.

A. Threshold of detection experiments

The results have confirmed the findings of the first report that the floor reflection will contribute on an individual basis to the timbre of a noise signal. For a speech signal none of the investigated reflections have been found to contribute on an individual basis to the timbre.

The introduced filtering of the individual transfer functions has a significant effect for reflections where spectral changes occur in the mid- and high-frequency ranges. Further experiments suggest that the threshold of detection is determined by the spectral changes in a dominant frequency range of 500 Hz–2 kHz.

The level of the reverberant field relative to the direct sound and early reflections has limited influence on the threshold values for ratios within the range 0.5–5 dB.

B. Just noticeable difference experiments

The results have confirmed the findings of the first report that an increase in the level of individual reflections for a noise signal is most likely to be audible for the first-order floor reflection, and for refections from the wall to the left of the listener. The present experiment has further shown that the first-order reflection from the wall behind the listener also belongs to this group. For a speech signal only the first-order floor reflection is most likely to produce an audible effect.

C. General

The findings discussed above suggest that the TD values for both noise and speech signals reported in the first report, define the perceptually most sensitive situation. Reflections where the mid- and high-frequency parts of the spectrum are attenuated by introducing the directivity of the loudspeaker and absorption of the room surfaces will have higher TD values.

The TD and jnd values for the speech signal are at a level where it cannot be excluded that the detection cue at threshold has been loudness and not timbre.

The results presented in this and the first report, have indicated that only certain reflections are likely to influence the timbre of the sound field on an individual basis. It is, however, very important to note, for future utilization of the results, that only *thresholds* have been measured. This means that the results cannot be used to predict how changes in those reflections will influence the timbral *quality* of the reproduction. Recently Walker[23] published results on a new type of control room design where the TD values presented here have been used as a design guide. Walker[24] has further discussions on the new design, including its effect on timbre of reproduction.

ACKNOWLEDGMENTS

The project team included the following people on a permanent basis. L. R. Fincham, Dr. R. H. Small, A. Jones, R. I. Harris, and R. Wilson all from KEF Audio Ltd., V. Hansen, G. Munch, and K. B. Christensen all from Bang & Olufsen A/S, and K. Rasmussen, O. Juhl Pedersen, N. Farver, and S. Bech all from The Acoustics Laboratory, J. Rasmussen, Å. Sonesson, H. Hansen from The Acoustics Laboratory are thanked for the never ending technical support. S. K. Pramanik is thanked for helpful comments on earlier versions of this paper. The entire group of subjects and experimenters are thanked for their enthusiasm and support throughout the project. Finally, the two reviewers are thanked for helpful comments.

[1] S. Bech, ''Timbral aspects of reproduced sound in small rooms. I,'' J. Acoust. Soc. Am. **97**, 1717–1726 (1995).

[2] L. Cremer and H. A. Müller (translated by T. J. Schultz), *Principles and Applications of Room Acoustics* (Applied Science, New York, 1982), Vol. 1, pp. 17–19.

[3] J. H. Rindel, ''Modeling the Angle-dependent Pressure Reflection Factor,'' Appl. Acoust. **38**, 223–234 (1993).

[4] Note that Rindel[3] concludes that the diffuse field coefficient should be set equal to the absorption at 55 degrees angle of incidence. The 55 degrees was found in experiments conducted after the present. The maximum difference is 0.9 dB for 100 and 125 Hz for reflection No. 5. For all other situations the difference is smaller than 0.5 dB.

[5] L. R. Fincham and R. H. Small, ''The application of digital signal processing to large scale simulation of room acoustics. Part I—Signal processing requirements in the Archimedes project,'' 90th Convention of The Audio Engineering Society, Paris, France, 19–22 February, Preprint 3055 (1991).

[6] D. M. Brookes, R. I. Harris, and R. J. Wilson, ''The application of digital signal processing to large scale simulation of room acoustics. Part III—DSP engine hardware topology and control software for multichannel simulation,'' 90th Convention of The Audio Engineering Society, Paris, France, 19–22 February, Preprint 3057 (1991).

[7] K. B. Christensen, ''The application of digital signal processing to large scale simulation of room acoustics. Part II—Frequency response modeling and optimization software for a multichannel DSP engine,'' J. Audio Eng. Soc. **40**, 260–276 (1992).

[8] S. Bech, ''Training of subjects for auditory experiments,'' Acta Acust. **1**, 89–99 (1994).

[9] ANSI S1.1-1960, ''Acoustical Terminology'' (American National Standards Institute, New York, 1960).

[10] The limitation in number of reflections investigated was caused by time constraints, and the selection of reflections was based on the TD's reported in Bech.[1] Those values can roughly be divided into three (four) groups according to level: group 1 includes reflections 1, 2, 3, and 4, group 2 includes reflections 5, 8, 9, and 12 and group 3 reflections 13–17. Reflections 6, 7, 10, and 11 are in a group of their own due to the mutual masking effects.[1] The reflections were selected to represent all three (four) groups. However, given the current level of understanding of the results, it is likely that reflections 3, 8, and 12 would have been included for further investigations.

[11] H. Levitt, ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477 (1971).

[12] S. E. Olive and F. E. Toole, ''The detection of reflections in typical rooms,'' J. Audio Eng. Soc. **37**, 539–553 (1989).

[13] F. A. Bilsen and R. J. Ritsma, ''Repetition Pitch and Its Implication for Hearing Theory,'' Acustica **22**, 63–73 (1969/70).

[14] Bilsen and Ritsma[13] only showed the existence of a spectral dominant range for the generation of repetition pitch of comb-filtered noise. However, Bilsen and Ritsma[18] also investigated timbre or coloration and found that the threshold of perceptibility of pitch and coloration of comb-filtered signals are similar and both have maximum for pitch values around 200 Hz corresponding to a dominant frequency range around 800 Hz.

[15] G. von Bismark, ''Timbre of steady sound: A factorial investigation of its verbal attributes,'' Acustica **30**, 146–159 (1974).

[16] G. von Bismark, ''Sharpness as an attribute of the timbre of steady sounds,'' Acustica **30**, 159–172 (1974).

[17] D. M. Green, *Profile Analysis Auditory Intensity Discrimination* (Oxford U.P., New York, 1988).

[18] F. A. Bilsen and R. J. Ritsma, ''Some Parameters Influencing the Perceptibility of Pitch,'' J. Acoust. Soc. Am. **47**, 469–475 (1970).

[19] The jnd values were not measured in the modified setup without filtering, so it is not possible to directly compare jnd values based on sound fields with and without filtering. However, given the following assumptions, it is possible to use the results from Bech[1] to represent the jnd values that would have been obtained in the modified setup without filtering: Previous discussions, and in the present Sec. V have assumed, and been supported by the results, that the timbral difference at the TD or jnd is determined by a spectral difference. Thus the TD's or jnd's obtained previously should be similar to those of the modified setup without filtering if the spectra of the two complete sound fields are similar. The spectral differences between the two complete sound fields for a broadband pink noise signal were evenly distributed over the spectrum and in the range of ± 2.5 dB for the frequency range 200 Hz–20 kHz. Thus to test above assumption, TD values were measured for the two sound fields for reflections 1, 5, 7, 9, 13, and 17 for noise and reflections 1, 7, 13, and 17 for speech. The results showed that only the TD's for reflections 13 and 17 were significantly different for the noise signal. It was not possible to find any plausible explanation for the difference in TD's for reflections 13 and 17. However, the results in general were taken as supportive of a similarity, in terms of TD and jnd values, between the two sound fields. This result further supports the discussion in Sec. V A 4, that the reverberant field only has a limited influence on the TD and jnd values for timbre, but a very significant influence on the spatial properties, and thereby the perceived reality of the simulation. Note that the threshold values based on the first report, shown in Figs. 8 and 9, have been calculated for the same four (noise) or three (speech) subjects as used for the filtered threshold values. The results shown in Fig. 5 in Bech[1] are based on all subjects, and will therefore differ from those shown here. Also note that only jnd values for reflections 1, 7, 13, and 17 for the speech signal, can be found in the first report.

[20] K. P. Tschopp and T. Beckenbauer, ''The Level Difference Limen of Test Materials used in Speech Audiometry,'' Acustica **75**, 173–183 (1991).

[21] S. Bech, ''Perception of Timbre of Reproduced Sound in Small Rooms: Influence of Room and Loudspeaker Position,'' J. Audio Eng. Soc. **42**, 999–1007 (1994).

[22] It is assumed that the absorption characteristics of the floor and ceiling are constant for the main parts of the ceiling and floor area and that the height of the loudspeaker and the ear level of the listener are constant.

[23] R. Walker, ''A new approach to the design of control room acoustics for stereophony,'' 94th Convention of The Audio Engineering Society, Berlin, Paper G1-1 (Preprint No. 3543) (1993).

[24] R. Walker, ''Early reflections in studio control rooms: the results from the first controlled Image Design installations,'' 96th Convention of The Audio Engineering Society, Amsterdam, Paper P12-6 (Preprint No. 3853) (1994).